dbGaP Study Release Notes

Release Notes for NHLBI TOPMed WGS GALAI, phs001542.v2.p1
"*NHLBI TOPMed: Genetics of Asthma in Latino Americans (GALA)*"

For any questions or comments, please contact: dbgap-help@ncbi.nlm.nih.gov.

December	17, 2019	Version 1 Data set release date
October	15, 2021	Version 2 Data set release date

2021-10-15
Version 2 Data set release for NHLBI TOPMed WGS GALAI now available

This release includes updated phenotype tables and the addition of Freeze 9 whole genome sequences (WGS) and corresponding VCFs. Please refer to the latest study configuration report for a detailed description of each download component.

Consent group 2 (c2): Disease-Specific (Lung Diseases, IRB, COL) (DS-LD-IRB-COL)

Data Type	subjects	samples
Phenotype	1024	1011
Seq_DNA_SNP_CNV (VCFs)	963	964
WGS	963	964

For a description of SAMPLE_USE terms, please see:
https://www.ncbi.nlm.nih.gov/projects/gap/submission/GetSampleUseTypes.cgi

Study and Phenotype Updates

1. **New Study Accession**

   NHLBI TOPMed WGS GALAI version 1 phs001542.v1.p1 has been updated to version 2. The dbGaP accession for the current phenotype data is **phs001542.v2.p1**. The participant number (p#) has not changed in version 2. No new subjects have been added to the study. However, samples have been retired and one sample has been remapped to a different subject.

2. **Updated Datasets (n=4 datasets)**

pht	version	Dataset Name
9642	2	TOPMed_WGS_GALAI_Subject
9643	2	TOPMed_WGS_GALAI_Sample
9644	2	TOPMed_WGS_GALAI_Subject_Phenotypes
9645	2	TOPMed_WGS_GALAI_Sample_Attributes

3. **New Variables (n=1 variable)**

pht	pht version	Dataset Name	phv	Variable Name
9642	2	TOPMed_WGS_GALAI_Subject	493821	SEX

4. **Retired Variables (n=1 variable)**

pht	Dataset Name	phv	version	Variable Name
9642	TOPMed_WGS_GALAI_Subject	420947	1	AFFECTION_STATUS

dbGaP Study Release Notes

Molecular Data Updates

dbGaP QC steps for this release consist of checks for consistency of subject and sample IDs in phenotype and genotype components.

1. For samples and marker/enrichment-procedure info, see download components:
    a. phg001668.v1.TOPMed_WGS_GALAI_v2_frz9.sample-info.MULTI.tar.gz
2. Genotypes are available in a matrix format as multi-sample vcf file(s) packed within download component(s) marked as genotype-calls-vcf. Integrity of submitted vcf files and their compatibility with PSEQ are routinely checked. Components may be divided by platform and/or population.
    a. phg001668.v1.TOPMed_WGS_GALAI_v2_frz9.genotype-calls-vcf.WGS_markerset_grc38.c2.DS-LD-IRB-COL.tar.gz

phg001303.v1	Freeze 8
phg001668.v1	Freeze 9

Authorized Access (Individual Level Data)

Individual level data are available for download through the dbGaP Authorized Access System upon approval of the Data Access Request (DAR):

- https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login

Public FTP site (Summary Level Data Only)

All data tables, data dictionaries, and documents will be housed under one directory for ease of downloading. The data_dict filenames have an added study version number (phs#.v#) and deleted participant set number (p#) from the table accession (pht#.v#). The var_report filenames have an added study version number (phs#.v#). In the var_report files, variables contain version numbers (phv#.v#) and summaries were created for each consent group (c#). These FTP files are available at:

- https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs001542/phs001542.v2.p1

2019-12-17

Version 1 Data set release for NHLBI TOPMed WGS GALAI now available

This release includes the addition of Freeze 8 whole genome sequences (WGS) brokered through the Sequence Read Archive (SRA), and VCFs derived from WGS. Please refer to the latest study configuration report for a detailed description of each download component.

Consent group 2 (c2): Disease-Specific (Lung Diseases, IRB, COL) (DS-LD-IRB-COL)

Data Type	subjects	samples
Phenotype	1024	1024
Seq_DNA_SNP_CNV (VCFs)	931	931
WGS*	931	931

*These data are brokered through the Sequence Read Archive (SRA). Please see Authorized Access instructions below.
For a description of non-SRA SAMPLE_USE terms, please see:
https://www.ncbi.nlm.nih.gov/projects/gap/submission/GetSampleUseTypes.cgi

Molecular Data

dbGaP QC steps for this release consist of checks for consistency of subject and sample IDs in phenotype and genotype components.

1. For samples and marker/enrichment-procedure info, see download components:
    a. phg001303.v1.TOPMed_WGS_GALAI.sample-info.MULTI.tar.gz
    b. phg001303.v1.TOPMed_WGS_GALAI.marker-info.MULTI.tar.gz
2. Genotypes are available in a matrix format as multi-sample vcf file(s) packed within download component(s) marked as genotype-calls-vcf. Integrity of submitted vcf files and their compatibility with PSEQ are routinely checked. Components may be divided by platform and/or population.
    a. phg001303.v1.TOPMed_WGS_GALAI.genotype-calls-vcf.WGS_markerset_grc38.c2.DS-LD-IRB-COL.tar.gz

Authorized Access (Individual Level Data and SRA Data)

Individual level data and Sequence Read Archive (SRA) data are available for download through the dbGaP Authorized Access System upon approval of the Data Access Request (DAR):

- https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login

Public FTP site (Summary Level Data Only)

All data tables, data dictionaries, and documents will be housed under one directory for ease of downloading. The data_dict filenames have an added study version number (phs#.v#) and deleted participant set number (p#) from the table accession (pht#.v#). The var_report filenames have an added study version number (phs#.v#). In the var_report files, variables contain version numbers (phv#.v#) and summaries were created for each consent group (c#). These FTP files are available at:

- https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs001542/phs001542.v1.p1